AD-A069 566 CLEMSON UNIV S C DEPT OF MATHEMATICAL SCIENCES F/G 12/1

A MONTE CARLO COMPARISON OF SOME RIDGE AND OTHER BIASED ESTIMAT--ETC(U)

JAN 78 A MITRA, R F LING N00014-75-C-0451

UNCLASSIFIED N-94 NL

| OF |

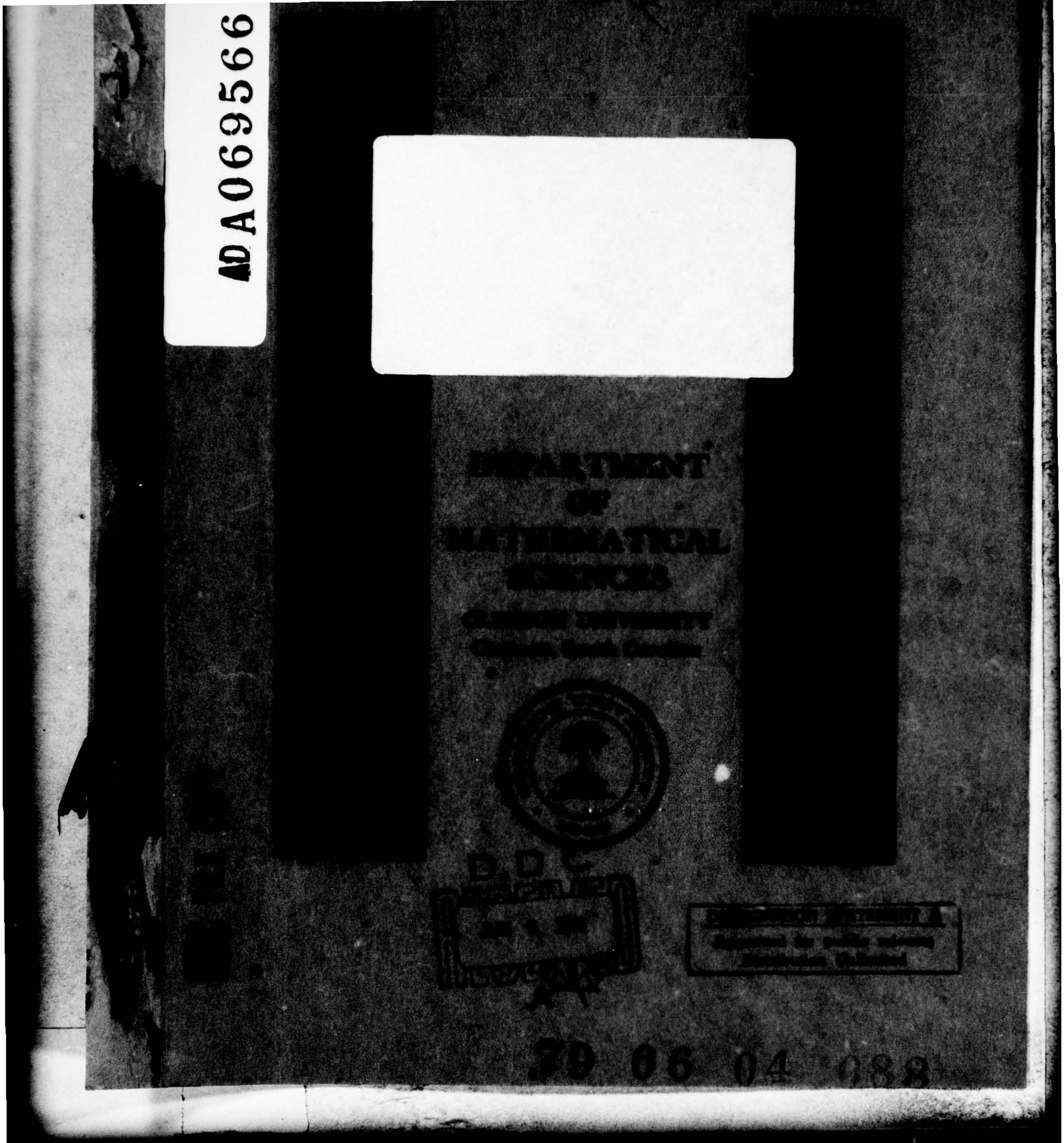
AD A069566

END

DATE FILMED

7-79

DDC



AD A069566

DEPARTMENT OF MATHEMATICAL SCIENCES

D D C

79 06 04 088

$k \leq 1$ was quite arbitrary, however. We considered two optimal ridge estimates: (1) The (global) optimal, denoted by $\hat{\beta}(k^*)$, where k^* minimizes $L(k)$ for $-\infty < k < \infty$; and (2) the positive-part optimal, denoted by $\hat{\beta}(k_+^*)$, where k_+^* minimizes $L(k)$ for $0 \leq k$. Dempster, Schatzoff, and Wermuth (1977) also used the notion of an optimal ridge estimate. However, the value of k for their optimal is determined from an expression that minimizes

(6)

A MONTE CARLO COMPARISON OF SOME RIDGE AND OTHER BIASED ESTIMATORS

BY

(10)

AMITAVA MITRA AND ROBERT F. LING

(14) N-94, TR-272

(9) TECHNICAL REPORT 272

(12) 36p.

(11) JAN[redacted]973

N-94

SUPPORTED IN PART BY THE OFFICE OF NAVAL RESEARCH
CONTRACT N00014--75-C-0451 TASK NR 042 - 271

(15) 407 183

1. INTRODUCTION

Consider the standard model for multiple linear regression

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}, \qquad (1.1)$$

where Y is an $n \times 1$ vector of observations on the dependent variable,

X is an $n \times p$ non-stochastic matrix of rank p,

$\underline{\beta}$ is a $p \times 1$ unknown vector of regression coefficients,

and $\underline{\varepsilon}$ is an $n \times 1$ random vector of errors with $E(\underline{\varepsilon}) = \underline{0}$ and

$$E(\underline{\varepsilon\varepsilon}') = \sigma^2 I.$$

The ordinary least squares (OLS) estimate of $\underline{\beta}$ is given by

$$\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{Y}. \qquad (1.2)$$

It is well-known that in the presence of multicollinearity condition among the independent variables, the OLS estimate $\hat{\underline{\beta}}$ is often very unstable and has, among its undesirable properties, a large value for its mean squared error (MSE).

Hoerl and Kennard (1970 a, b) introduced a class of estimators, termed <u>ridge</u> estimators, defined by

$$\hat{\underline{\beta}}(k) = (X'X + kI)^{-1}X'\underline{Y}, \ k \geq 0. \qquad (1.3)$$

Originally, (1.3) was used in conjunction with a "ridge trace" to find a constant k which would stabilize the components of the vector $\hat{\underline{\beta}}(k)$. In so doing, it was found that the stabilized

$\underline{\hat{\beta}}(k)$ often resulted in a reduction of MSE as well, as compared to the OLS $\underline{\hat{\beta}}$.

Subsequent literature on ridge estimation focussed attention on the MSE properties of $\underline{\hat{\beta}}(\hat{k})$, where $\hat{k}$'s (which are functions of the data $\underline{X}$ and $\underline{Y}$) have been proposed and studied. For a review of the ridge regression literature and some recent simulation results, see, e.g., (Hocking 1976, Dempster, Schatzoff, and Wermuth 1977, and Gunst and Mason 1977).

In spite of the fact that the large amount of work on ridge regression deals almost exclusively with the MSE property or performance of various ridge estimators, MSE considerations should remain secondary in importance in the ridge regression context although the MSE criterion is an important estimation criterion in its own right. The two main reasons for this perspective are: (1) The primary intent of a ridge analysis is to remove some undesirable effects of multicollinearity - this can be accomplished either by keeping all of the original independent variables and attempt to minimize the MSE using (1.3) as the form of the point estimate, or by removing one or more independent variables which cause the multicollinearity condition, by using ridge regression as a tool for detecting such variables. (2) if MSE were the primary concern of the analysis, there would be considerably less incentive to restrict one's attention to the class of ridge estimators, in which very little theoretical results are known; whereas several other classes of estimators (relatives of the James-Stein (1961) estimator) have been proven to dominate the OLS estimate in MSE.

For the above reasons, we carried out a simulation experiment (involving several ridge estimators) which was different from previous studies by others in several respects. First, we did not attempt to do a comparative study involving a large number of estimators as was done by Dempster, Schatzoff and Wermuth (1977), nor did we restrict our attention to ridge estimators alone. Instead, we selected several ridge estimators that have previously been reported to have good MSE properties, as well as several estimators not in the ridge-class that seem to hold promise for theoretical reasons. Second, we controlled the relevant factors and parameters of the problem over a more comprehensive region than those considered in previously reported studies. Third, we studied in detail the empirical sampling distributions of the squared loss incurred by each method over various combinations of a factorial design. Fourth, several new quantities of calibration were used, in addition to the empirical MSE, in assessing and comparing the performance of various estimators.

Within the scope of the present study, we found, with rare exceptions (in near orthogonal cases of the design matrix), all of the estimators to yield better empirical MSE than that of the OLS estimates, and often by substantial margins especially under conditions of high multicollinearity among the independent variables. On a relative basis, several different estimators excel in different regions of the control-factor space but none was found to be the best in all the regions.

On the other hand, two of the best estimators for highly multicollinear data were among the worst when the design matrix X was nearly orthogonal. Several of the estimators have been proven theoretically to dominate the OLS $\hat{\underline{\beta}}$ in MSE. However, none of the existing theoretical results enables one to assess the magnitudes of improvement over our control-factor space. We found in many instances the estimators whose theoretical MSE's are unknown performed considerably better (in empirical MSE) than those in which theoretical MSE results exist.

2. ESTIMATORS CONSIDERED IN THIS STUDY

Several biased estimators were considered in the study along with the OLS estimator. The selection process was necessarily subjective. One of the factors influencing our choice was our a priori judgement of the potential of the estimators guided by studies reported in the most recent literature. Moreover, the fact that some simulation studies had been done on certain individual estimators motivated us to analyze the relative performance of these estimators when stu-studied collectively. This, we felt, would provide some insight with respect to certain measures of performance of the dominance of certain estimators over others in some ranges of the factor space, which would enable us to develop a guideline for choosing estimators for a particular problem.

The estimators that we chose may be grouped into two categories: single-parameter or multi-parameter families. Within these groups we have considered some ridge estimators and some other estimators that were originally proposed in the context of estimation of the mean of a multivariate normal distribution. James and Stein (1961) estimator and the Baranchik (1970) class of estimators are examples of the latter type. The equivalence of ridge-type estimators and Baranchik-type estimators have been shown for the orthogonal case ($X'X = I$) and may be found in (Mitra 1977). Two new classes of ridge estimators were derived from the Baranchik class of estimators.

The Baranchik class of estimators, for the estimation of the mean of a multivariate normal distribution is of the form

$$\delta_i = (1 - \frac{p-2}{s}\tau(s))x_i, \tag{2.1}$$

where $s = \underline{x}'\underline{x}$, $\underline{x} = (x_1, x_2, \ldots x_p)$, and $\tau(s)$ is a function of s. For each of p parameters $\theta_1, \theta_2, \ldots\theta_p$ we observe an independent normal variate $X_i \sim N(\theta_i, 1)$. Baranchik (1970) found sufficient conditions for $\tau(s)$ to guarantee the estimator to have smaller risk then the usual estimator $\underline{x}$. Efron and Morris (1976) found necessary and sufficient conditions for estimators of the form (2.1) to dominate $\underline{x}$ in risk. They considered the case when $x_i \sim N(\theta_i, \sigma^2)$ where σ^2 may be unknown. In their formulation, s in (2.1) is replaced by F, where $F = \frac{\underline{x}'\underline{x}(n-p+2)}{\hat{\sigma}^2(n-p)}$.
The two new classes of estimators, labeled as Mitra 1(M1) and Mitra 2 (M2) are special cases of the above form, and will therefore dominate the OLS estimator in MSE in the orthogonal case. For the Mitra 1 estimator (1977), the ridge parameter k in (1.3) takes on the form

$$\hat{k} = \frac{(p-2)(t - \frac{c_1}{c_2+F})}{F - (p-2)(t - \frac{c_1}{c_2+F})}, \text{ where } F = \frac{\hat{\underline{\beta}}'\hat{\underline{\beta}}(n-p+2)}{(n-p)\hat{\sigma}^2},$$

where $\hat{\underline{\beta}}$ is the OLS estimator and $\hat{\sigma}^2$ is an unbiased estimator of σ^2.

For the simulation, the values of the parameters were chosen to be t=1, c_1=1, and c_2=2 yielding

$$\hat{k} = \frac{(p-2)}{F(F+2)/(F+1) - (p-2)} \tag{2.2}$$

For the Mitra 2 estimator the ridge parameter is of the form

$$\hat{k} = \frac{(p-2)[t - c_1 e^{-c_2 F}]}{F - (p-2)[t - c_1 e^{-c_2 F}]} .$$

For the simulation, the parameter values were chosen to be $t=1$, $c_1=1$, $c_2=1$, yielding

$$\hat{k} = \frac{(p-2)(1-e^{-F})}{F - (p-2)(1-e^{-F})} . \tag{2.3}$$

A multi-parameter or generalized ridge estimator takes the form

$$\hat{\beta}(K) = (X'X + KI)^{-1}X'Y, \tag{2.4}$$

where $$K = \mathrm{diag}(k_1, k_2, \ldots, k_p).$$

Table 1 gives a listing of the estimators that were used in the simulation, along with their references. The reader is referred to these references for explicit expressions of these estimators, which are omitted from this paper.

HK and GM are special cases of (2.4) while B, B+, V, and V+ are of the form (2.4) but K is not necessarily diagonal.

3. SIMULATION DESIGN

Many simulation studies on the performance of ridge and other biased estimators have appeared in the recent literature. Among these are the studies of Gunst and Mason (1977), Dempster, Schatzoff and Wermuth (1977), Lawless and Wang (1976), Hoerl, Kennard and Baldwin (1975), McDonald and Galarneau (1975), to cite but a few. Our study differs from these and other simulation studies mainly in two respects. First, we considered the relevant parameters of the problem (dimensionality, degree of multicollinearity, etc.) over a more comprehensive range and combinations than previous studies. Second, we report the empirical performance of various estimators in greater detail, i.e., using several measures of performance in addition to the usually reported average squared-loss (empirical MSE).

3.1 Generation of X and $\underline{\beta}$.

In the ridge regression studies using simulations that have appeared to date, there does not appear to be any standard method of simulating data and parameters from the linear model $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$. In principle, the performance of ridge estimators depends on the design matrix X only through the eigenvalues of the matrix X'X, as was pointed out by Efron and Morris (1977) p. 92), and that X, $\underline{\varepsilon}$, or $\underline{Y}$ need not be actually generated, as was the case in the study of Dempster, Schatzoff and Wermuth (1977). In some studies, e.g., (Hoerl, Kennard and Baldwin 1975) particular sets of real data from other published sources

were used for X. We chose to follow the procedure used and described by McDonald and Galarneau (1975, p. 409) to generate the X's and $\underline{\beta}$'s in this study because the procedure provides a reasonable method of choosing an X (and the most and least favorable $\underline{\beta}$'s) in any given dimension with a specified multi-collinearity structure. Basically, for each given dimension p and a correlation coefficient ρ, the elements of X (x_{ij}, $i=1, \ldots, 100$; $j=1, \ldots, p$.) are assumed to have the intraclass correlation structure (theoretically):

$$\begin{pmatrix} 1 & \rho \ldots & \rho \\ \rho & 1 \ldots & \rho \\ \vdots & & \\ \rho & \ldots & 1 \end{pmatrix} \tag{3.1}$$

The sample size of 100 was used (as was in McDonald and Galarneau 1975) so that the sample correlation matrix of X closely resembles the form (3.1). For each X, two sets of $\underline{\beta}$ are generated, of unit length, corresponding in some sense to the "most favorable" (where $\underline{\beta}$ is the normalized eigenvector corresponding to the largest eigenvalue of the X'X matrix) and the "least favorable" (normalized eigenvector corresponding to the smallest eigenvalue of X'X) choices of $\underline{\beta}$ for that X. See (McDonald and Galarneau 1975, p. 409).

3.2 Choice of Parameters and Method of Replication.

The dimensionality of X (the number of independent variables), denoted by p, were taken to be 3, 6, and 10, corresponding to what we considered to be low, moderate, and high dimensions.

For each p, six values of ρ were chosen (as a function of the multicollinearity index α described in the following subsection 3.3) ranging from $\rho = 0$ to $\rho > 0.99$. For each combination of p and ρ, an X matrix and two coefficient vectors $\underline{\beta}$ were generated as described in the previous subsection. For each of these combinations of (p, ρ, X, $\underline{\beta}$), random error vectors $\underline{\varepsilon}$ (and hence $\underline{Y}$) were replicated 100 times from each of 5 normal distributions $N(\underline{0}, \sigma^2 I)$. The values of σ^2 considered in this study were 0.001, 0.01, 0.1, 0.2, and 1.0, corresponding to the relative magnitudes of σ^2 to $\underline{\beta}$ (which has unit length).

3.3 <u>Index of Multicollinearity</u> α.

Given the correlation structure (3.1) of X, McDonald and Galarneau (1975) used ρ as their measure of multicollinearity in X. We introduced an index α, which is a function of ρ and p as our measure of multicollinearity, derived from the following heuristics:

In general, a reasonable measure of multicollinearity is

$$\delta = \sum_{i=1}^{p} 1/\lambda_i,$$

where λ_i, i=1, ..., p are the eigenvalues of X'X. For X'X of the form (3.1),

$$\delta = (p-1)/(1-\rho) + 1/(1 + (p-1)\rho), \tag{3.2}$$

so that $p \leq \delta < \infty$. In particular, if $\rho = 0$, then $\delta = p$. We therefore introduced a "normalized" measure

$$\alpha = \delta/p \tag{3.3}$$

so that $\alpha = 1$ when $\rho = 0$ for all p, and $1 \leq \alpha < \infty$. It can be easily seen from (3.3) and (3.2) that given p and α, the corresponding $\rho \geq 0$ is

$$\rho = [(p-2)(\alpha-1)+((p-2)^2(\alpha-1)^2+4\alpha(\alpha-1)(p-1))^{\frac{1}{2}}]/(2\alpha(p-1)). \quad (3.4)$$

In this study, we chose the values of α to be 1, 2, 5, 10, 50, and 100, which correspond to the values of ρ shown in Table 2.

4. SIMULATION RESULTS

Several performance measures were used to study the relative performance of the fourteen estimators (including OLS).

4.1 Comparisons to optimal values.

Given $\underline{\beta}$ (which is the unknown coefficient vector to be estimated, but known in a simulation study) and a realization of the error vector $\underline{\varepsilon}$, values of k can be determined such that the squared loss

$$L(k) = (\hat{\beta}(k)-\beta)'(\hat{\beta}(k)-\beta),$$

where $\hat{\beta}(k)$ is of the form (1.3), is minimized over various intervals for k. These minimizing values of k can be said to yield optimal single-parameter ridge estimates of $\underline{\beta}$. Such optimal estimates are of course not realizable in practice because they depend on the particular realizations of $\underline{\varepsilon}$ as well as the value of true $\underline{\beta}$. However, they are realizable in a simulation study and serve as useful quantities of calibration because they yield absolute lower bounds for L(k) over the entire class of single-parameter ridge estimators, irrespective of how k is determined from the data.

McDonald and Galarneau (1975) considered such optimal ridge estimates, for k restricted to the unit interval [0, 1]. The restriction $k \geq 0$ was natural because the restriction is generally included in the definition of a ridge estimator. The restriction

$k \leq 1$ was quite arbitrary, however. We considered two optimal ridge estimates: (1) The (global) optimal, denoted by $\hat{\beta}(k^*)$, where k^* minimizes $L(k)$ for $-\infty < k < \infty$; and (2) the positive-part optimal, denoted by $\hat{\beta}(k_+^*)$, where k_+^* minimizes $L(k)$ for $0 \leq k$. Dempster, Schatzoff, and Wermuth (1977) also used the notion of an optimal ridge estimate. However, the value of k for their optimal is determined from an expression that minimizes the MSE (using estimates of the OLS) instead of minimizing the squared loss for each realization of $\underline{\varepsilon}$.

Three indices, $L(\hat{k})/L(k^*)-1$, $L(\hat{k})/L(k_+^*)-1$ and $1 - L(\hat{k})/L(0)$, were used to compare the performance of an estimator with that of the optimal ridge, positive part optimal ridge and the OLS estimator respectively for each sample. A sample of such results is shown in Table 3. From this table we find that for this sample, methods M, HKB, JS+, B+, and B do worse than the OLS in squared loss. As far as performance with respect to the optimal ridge estimate is concerned, we note from the same table that M was 2.374 times worse than it. On the other hand, HK did better than the single parameter optimal ridge estimator, which is possible since HK is a multi-parameter ridge estimator. For this sample, $k^* > 0$, so that $k^* = k_+^*$.

These optimal ridge estimates serve a useful purpose in MSE comparisons also. Along with the empirical MSE of the different estimators, found over the replications, the empirical MSE of the optimal ridge estimates are found in the simulation. One such result is shown in Table 4. Such comparisons will give us

a measure of how much of a reduction in MSE is possible, even though not attained. In particular, from Table 4, for $\alpha = 1$, we find HKB to be the best with a MSE of 1.20. The OLS has an empirical MSE of 2.89. The OLS (Expected) row represents the theoretical MSE for the OLS as found from $\sigma^2 \sum_{i=1}^{p} 1/\lambda_i$. Hence, the best estimator in this case (HKB) has MSE a little less than one half (0.415) of that of the OLS. On the other hand, in comparison to the lower bound of the MSE, the MSE of HKB is 2.4 times larger than the MSE of the optimal ridge estimator.

From Table 4, we may also note the relation between the MSE of the optimal and the MSE of the positive part optimal ridge estimator. As the level of multicollinearity increases, even though the same trend in the MSE's is observed, the ratio of the MSE of the positive part optimal to the MSE of the optimal ridge estimator is found to increase from 1.1 to 17.0. This implies that much reduction in L(k) is potentially achievable by not restricting k to be nonnegative, at high levels of multicollinearity. However, when $\underline{\beta}$ is least favorable, it was found that there was not a large difference in the MSE's of the optimal and positive part optimal ridge estimates. Other similar tables of results may be found in (Mitra 1977).

4.2 Frequency Comparisons.

The empirical MSE is a useful measure of performance. However, other measures may be more informative, especially if the distribution of $L(\hat{k})$ is highly skewed. The sampling distributions and certain fractiles of $L(\hat{k})$ of the various estimators were

studied in Mitra (1977) where we found some cases in which $L(\hat{k})$ is smaller than its OLS counterpart $L(0)$ a large percent of the time while the MSE of $\hat{\beta}(0)$ is smaller than that of $\hat{\beta}(k)$. We present a squared loss comparison to the OLS estimates in Table 5, which shows the approximate percentage of times estimators have smaller squared loss than that of OLS. McDonald and Galarneau (1975) carried out a similar analysis. However, their study did not consider all of the estimators that we have simulated. Moreover, the effect of the degree of multicollinearity and $\underline{\beta}$ can be observed from our results. From Table 5 we find that all of the estimators, with the exception MG, outperform the OLS estimator a large fraction of the time. For high multicollinearity, except for MG, the ridge estimators have less loss almost 100% of the time. The effect of α can be seen as we notice that for low multicollinearity, the estimators have a smaller loss than that of the OLS a smaller proportion of times than when α is high. The effect of $\underline{\beta}$ can also be observed from Table 5, and satisfies our intuitive conclusions. With the exception of B and B+, all the estimators perform better when $\underline{\beta}$ is most favorable than when $\underline{\beta}$ is least favorable.

4.3 MSE Comparisons

One of the common measures of performance is the mean square error given by

$$MSE = E[L(k)] = E[(\hat{\underline{\beta}}(k)-\underline{\beta})'(\hat{\underline{\beta}}(k)-\underline{\beta})]. \qquad (4.1)$$

For each method and each combination of the choice of parameters p, σ^2, α, and $\underline{\beta}$, the empirical MSE is computed over one hundred replications.

Following Gunst and Mason (1977), an analysis of variance procedure was adopted to determine the effects of the controlled parameters. The logarithm of the empirical mean square error was considered in the analysis. The main effects of α, p, and σ^2 are highly significant for all the methods. The effect of $\underline{\beta}$ is also quite significant except for the OLS and V_+ estimators. Also, the two-way interactions have significant effects in most of the estimators considered.

A sample table of results showing the empirical MSE of the different estimators including the OLS and the optimal ridge and positive part optimal ridge estimator may be found in Table 4. Other similar tables showing MSE results for the different parameter combinations are omitted from this paper and can be found in (Mitra 1977).

The effect of α, a measure of the degree of multicollinearity, may be noticed from Table 4. As α increases, the MSE of the various estimators tends to increase, which seems intuitively expected. The relative performance of some of the estimators is very much affected by α too, as can be seen from Tables 7 and 8. The two most noticeable ones are GM and M which are among the "worst MSE" when α is small, but dramatically among the "best MSE" when α is large.

The effect of σ^2, the variance of the error term, also behaves as expected. As σ^2 becomes small, the MSE of the various estimators decreases. For very small values of σ^2, say $\sigma^2 = 0.001$, there is not much difference in the actual magnitudes of the MSE of the different estimators. This suggests that for small σ^2, one may use the OLS estimator and have MSE comparable to the ridge or other biased estimators.

The unknown coefficient vector $\underline{\beta}$ has an effect on the MSE of the optimal ridge estimates. When $\underline{\beta}$ is most favorable, then as α increases the MSE of the optimal ridge estimate tends to decrease. Hence the ratio of the MSE of any chosen estimator to the MSE of the optimal ridge estimator, increases as α increases, when $\underline{\beta}$ is most favorable. On the other hand, when $\underline{\beta}$ is least favorable, as α increases the MSE of the optimal ridge estimator increases, just as that of the other estimators. The effect of $\underline{\beta}$ with respect to the percentage of times an estimator does better than OLS in squared loss was observed in Section 4.2.

The effect of p, the number of independent variables, does not exhibit any general pattern of influence on the behavior of all the estimators. However, the behavior of some of the specific estimators seems to depend on p, e.g., in terms of MSE, B and B+ do well when p is large, and these estimators change from the "worst MSE" class to the "best MSE" class as p changes from small to large. The reverse holds for the estimator F.

4.4 Empirical Rank Minimax Analysis

Frequently one is interested in obtaining an estimator that is minimax, with respect to some measure of performance. A minimax estimator would ensure the user that the worst performance of it will be better than the worst performance of the other estimators. Using the empirical MSE as measure of performance, we found the maximum rank of the MSE's of the various estimators, over the choice of parameters of: $p=3, 6, 10$; $\alpha=1, 2, 5, 10, 50, 100$; $\sigma^2=1.0, 0.1$; and $\underline{\beta}$. Table 6 shows the maximum rank of the MSE's of the different estimators. We observe that from a minimax point of view, M2 and M1 as well as HKB and HK estimators seem most favorable, and the latter seems consistent with one of the findings of Dempster, Schatzoff and Wermuth (1977).

4.5 Tables of Guideline

Using the simulation results, we now devise, as a rule of thumb, a procedure to choose estimators that may be preferred as well as those that may be avoided, under the various combinations of the parameter conditions. For this purpose, we divide the level of multicollinearity into two parts, small α for $\alpha < 3$ and large α for $\alpha \geq 3$. The variance of the error term, σ^2, is partitioned into two levels, corresponding to regions, $\sigma^2 \leq 0.1$ and $\sigma^2 > 0.1$ respectively. Similarly the number of independent variables, p is divided into two parts. Under this structure, the MSE and the ranks associated with them for the various estimators were used to form the guideline tables. The average ranks of the MSE's of each of the fourteen simulated estimators (which

includes the OLS), under the partitioned structure so described, were found and used as a basis for choosing estimators that are in the "best MSE" class and those that are in the "worst MSE" class. Table 7 shows the guideline for single-parameter estimators while Table 8 depicts the guideline for multi-parameter estimators. The OLS estimator is not included in the table, since for our range of controlled parameters the OLS estimate would always fall within the "worst MSE" class.

Some general observations may be drawn from these guideline tables. Irrespective of the other parameters, for small α, GM and M are in the "worst MSE" class, while for large α they are in the "best MSE" class. B and B+ are among the best when p is large while M2, M1, HKB, and JS+ are among the best when α is small. In view of the marked dependence of the qualities of the various estimators on the parameter space of the problem, we believe Tables 7 and 8 will aid the user of ridge estimation methods in choosing a favorable method for his particular problem.

5. CONCLUDING REMARKS

In this study we considered the performance of some ridge and other biased estimators under a wide range and combination of controlled parameter values. For a given regression problem, α and p are known and σ^2 can be estimated, so that one may use the tables of guideline to select an estimation method for $\underline{\beta}$.

In view of the extensive nature of this study, we feel that our results, bolstered by supporting conclusions from other published simulation studies, enable us to recommend with confidence a method of choosing among a large class of estimators a small number of promising candidates, on the basis of certain characteristics of particular problems.

For each set of simulated data, the squared loss of the $\underline{\beta}$ estimate of each estimation method was compared to two common calibrations, the squared loss of the OLS $\hat{\underline{\beta}}$ and that of the optimal ridge $\hat{\underline{\beta}}(k^*)$. The former lets us guage the magnitude of squared loss improvement over the OLS estimate while the latter, being the absolute minimum squared loss for using a single-parameter ridge estimate for $\underline{\beta}$, enables us to obtain an empirical relative-efficiency of each method as well as observing whether any of the other estimators is capable of achieving a smaller squared loss (or a smaller average) than the minimum loss that could possibly be achieved by a member of the single-parameter ridge class of estimators.

APPENDIX

The computer programs used in this study were coded by A. Mitra. Computations were performed in double-precision on an IBM/370 model 165 machine, with programs coded in FORTRAN and compiled by WATFIV (Version I, level 5).

The reported simulation results were based on the use of subroutines RANDU and GAUSS from the IBM Scientific Subroutine Package (1970) for pseudorandom number generation. In addition, the statistical reliability of certain results (including Table 4) were verified by the use of better uniform and normal generators:

The uniform pseudorandom number generator was

$$X_{i+1} = 764{,}261{,}123\ X_i \pmod{2^{31} - 1},$$

which was reported by Hoaglin (1976) to have excellant spectral and lattice properties. Its spectral numbers are $C_2 = 1.94$, $C_3 = 2.10$, $C_4 = 2.58$, $C_5 = 4.06$, and $C_6 = 2.55$; and its lattice numbers L_i are less than 2, $i=2, \ldots, 6$. Random normal deviates were generated by applying the Box-Muller (1958) transformation to the uniform (0, 1) numbers produced by the Hoaglin generator described above.

1. Estimators Considered in Study

Single-Parameter:		
Ridge	HKB :	Hoerl, Kennard and Baldwin (1975, eq. 2.2)
	F :	Farebrother (1975, eq. 6)
	M :	Mallows (1973, p. 673), Farebrother (1975, eq. 11)
	MG :	McDonald and Galarneau (1975, Rule R_2)
	M1 :	(1.3) and (2.2)
	M2 :	(1.3) and (2.3)
Others	JS+ :	James and Stein (1961), Vinod (1976, eq. 13, p. 6)
Multi-Parameters:		
Ridge	HK :	Hoerl and Kennard (1970 a, p. 63)
	GM :	Guilkey and Murphy (1975, p. 770, DRE1, $\lambda_i < 0.1\lambda_{max}$)
Others	B, B+:	Bhattacharya (1966), Vinod (1976, eqs. 19, 21)
	V, V+:	Vinod (1976, eqs. 22, 23)

2. Correlations (ρ) for a Degree of Multicollinearity (α) for Various Dimensions (p)

Degree of Multicollinearity	Correlation (ρ) Dimension		
α	p=3	p=6	p=10
1	0.0	0.0	0.0
2	0.6404	0.5742	0.5462
5	0.8633	0.8322	0.8196
10	0.9325	0.9164	0.9099
50	0.9866	0.9833	0.9820
100	0.9933	0.9917	0.9910

3. Performance of Estimators with Respect to the OLS and Optimal Ridge Estimates for a Sample.

$p=3$, $\sigma^2=0.1$, $\alpha=2$, β=most favorable

Estimators	$1 - L(\hat{k})/L(0)$	$L(\hat{k})/L(k^*) - 1$	$L(\hat{k})/L(k_+^*) - 1$
GM	0.000	0.248	0.248
M	-0.902	1.374	1.374
HKB	-0.217	0.518	0.518
HK	0.527	-0.410	-0.410
F	0.196	0.004	0.004
M2	0.188	0.014	0.014
M1	0.196	0.003	0.003
MG	0.000	0.248	0.248
JS+	-0.179	0.472	0.472
V	0.029	0.212	0.212
V+	0.029	0.212	0.212
B+	-0.259	0.572	0.572
B	-0.259	0.572	0.572

4. Empirical Mean Square Errors (Standard Errors) and Their Ranks

p=3, σ^2=1.0, $\underline{\beta}$=most favorable

	α=1		α=2		α=5		α=10		α=50		α=100		Maximum
	MSE	Rank	MSE	Rank	MSE	Rank	MSE	Rank	MSE	Rank	MSE	Rank	Rank
Optimum	0.50		0.24		0.11		0.07		0.02		0.01		
Pos Part Optimum	0.55		0.39		0.33		0.31		0.16		0.17		
GM	2.89	12.5	6.56	12.5	1.81	1	6.56	1	22.57	1	60.25	1	12.5
	(0.20)		(0.63)		(0.64)		(2.11)		(10.08)		(20.92)		
M	6.91	14	10.86	14	4.01	2	18.30	3	46.78	2	102.79	2	14
	(2.08)		(6.20)		(0.56)		(6.60)		(10.39)		(21.09)		
HKB	1.20	1	2.23	1	4.24	3	11.39	2	46.83	3	103.19	3	3
	(0.09)		(0.33)		(0.59)		(1.78)		(10.42)		(21.11)		
HK	1.51	2	3.19	2	6.94	4	17.77	4	72.23	4	160.97	4	4
	(0.13)		(0.44)		(0.92)		(2.49)		(12.54)		(27.44)		
F	1.87	6	3.98	4	8.41	6	21.42	6	88.41	5	194.32	5	6
	(6.150)		(0.50)		(1.01)		(2.73)		(14.42)		(31.76)		
M2	1.74	3	3.83	3	8.29	5	21.39	5	88.99	6	195.78	6	6
	(0.15)		(0.50)		(1.01)		(2.74)		(14.48)		(31.91)		
M1	1.86	5	3.98	5	8.47	7	21.59	7	89.18	7	195.98	7	7
	(0.15)		(0.50)		(1.02)		(2.74)		(14.48)		(31.91)		
MG	2.05	7	4.25	6	9.50	8	23.69	8	99.44	8	228.38	8	8
	(0.15)		(0.46)		(0.83)		(2.30)		(13.01)		(27.74)		
JS+	1.81	4	4.46	7	9.91	9	25.51	9	114.65	9	240.33	9	9
	(0.15)		(0.50)		(1.06)		(3.02)		(15.25)		(32.97)		
V	2.61	10	6.08	9	14.10	11	34.69	10	150.71	10	322.55	10	11
	(0.19)		(0.60)		(1.28)		(3.37)		(17.08)		(37.98)		
V+	2.62	11	6.07	8	14.09	10	34.73	11	150.81	11	322.85	11	11
	(0.19)		(0.60)		(1.28)		(3.36)		(17.07)		(37.96)		
B+	2.19	8	6.26	10	15.11	13	37.04	13	161.40	12	349.02	12	13
	(0.16)		(0.61)		(1.37)		(3.43)		(17.50)		(39.72)		
B	2.21	9	6.47	11	15.10	12	37.03	12	161.42	13	349.03	13	13
	(0.16)		(0.60)		(1.37)		(3.44)		(17.50)		(39.72)		
OLS(Observed)	2.89	12.5	6.56	12.5	15.38	14	37.34	14	161.61	14	349.23	14	14
OLS (Expected)	3.32		6.76		17.46		35.47		180.70		362.65		

5. Approximate Percentage of Times Estimators Have Smaller Squared Loss than OLS.

		Low Multicollinearity $\alpha=1, 2$ $\underline{\beta}$ Most Favorable	$\underline{\beta}$ Least Favorable	High Multicollinearity $\alpha=5, 10, 50, 100$ $\underline{\beta}$ Most Favorable	$\underline{\beta}$ Least Favorable
Single-Parameter:					
Ridge:	HKB	96[a]	89	100	98
	F	98	91	100	98
	M	88	84	100	98
	MG	59	54	46	44
	M1	96	90	100	98
	M2	95	90	100	98
Others:	JS+	95	92	100	98
Multi-Parameter:					
Ridge:	HK	99	94	100	98
	GM	37	33	100	97
Others:	B	88	94	90	95
	B+	92	97	93	99
	V	99	85	99	95
	V+	100	85	99	95

[a]Average of the performance criterion (% of times loss of estimator < loss of OLS) over all combinations of parameters: $p=3, 6, 10$; $\sigma^2=1.0, 0.1$.

6. Maximum Rank of the Mean Squared Error of Estimators

Estimator	Maximum Rank of MSE[a]
M2	7
HKB	8
M1	9
HK	9
JS+	13
MG	13
F	13
B+	13
B	13
GM	14
M	14
V+	14
V	14

[a] Maximum Rank over all combinations of parameters: $p=3, 6, 10$; $\alpha=1, 2, 5, 10, 50, 100$; $\sigma^2=1.0, 0.1$; $\underline{\beta}$=most favorable and least favorable.

7. Guideline for Choosing Single-Parameter Estimators

	Low Multicollinearity (α=1,2)				High Multicollinearity (α=5,10,50,100)			
	Small variance ($\sigma^2 \leq 0.1$)		Large variance ($\sigma^2 > 0.1$)		Small variance ($\sigma^2 \leq 0.1$)		Large variance ($\sigma^2 > 0.1$)	
	p=3,6	p=10	p=3,6	p=10	p=3,6	p=10	p=3,6	p=10
"Best MSE"	HKB(3+)[a]	HKB(5)	JS+(3)	JS+(2+)	M(2)	M(2+)	M(2+)	M(3+)
		M2(6)	HKB(1+)	M2(4)				
		M1(6)	M2(3)	M1(3+)				
"Worst MSE"	M(12)	F(11+)	M(13)	F(12)	MG(9)	F(13)	MG(9+)	F(13)
		MG(10+)		MG(9)		MG(10+)		MG(10+)

[a]Numbers in parenthesis represent ranks of MSE averaged over the corresponding combination of parameters and $\underline{\beta}$ = most favorable and least favorable.

8. Guideline for Choosing Multi-Parameter Estimators

	Low Multicollinearity (α=1,2)				High Multicollinearity (α=5,10,50,100)			
	Small variance ($\sigma^2 \le 0.1$)		Large variance ($\sigma^2 > 0.1$)		Small variance ($\sigma^2 \le 0.1$)		Large variance ($\sigma^2 > 0.1$)	
	p=3,6	p=10	p=3,6	p=10	p=3,6	p=10	p=3,6	p=10
"Best MSE"	HK(2)[a]	B+(1+)	HK(4+)	B+(2)	GM(1+)	GM(1+)	GM(1)	GM(1+)
						B+(2+)		B+(2+)
"Worst MSE"	GM(11+)	GM(10)	GM(11+)	GM(10)	B(10)	V(10)	B(10+)	V(11)
	V(10+)	V(10+)	V(10+)	V(11)	B+(9)	V+(11)	B+(9)	V+(12)
	V+(10+)	V+(10)	V+(10+)	V+(11+)	V(11+)		V(11+)	
					V+(11+)		V+(12)	

[a]Numbers in parenthesis represent ranks of MSE averaged over the corresponding combination of parameters and $\underline{\beta}$ = most favorable and least favorable.

REFERENCES

Baranchik, A. J. (1970), "A Family of Minimax Estimators of the Mean of a Multivariate Normal Distribution," Annals of Mathematical Statistics, 41, 642-5.

Bhattacharya, P. K. (1966), "Estimating the Mean of a Multivariate Normal Population with General Quadratic Loss Function," Annals of Mathematical Statistics, 37, 1819-24.

Box, George E. P., and Muller, Mervin E. (1958), "A Note on the Generation of Normal Deviates," Annals of Mathematical Statistics, 29, 610-1.

Dempster, Arthur P., Schatzoff, Martin, and Wermuth, Nanny (1977), "A Simulation Study of Alternatives to Ordinary Least Squares," Journal of the American Statistical Association, 72, 77-91.

Efron, Bradley, and Morris, Carl (1976), "Families of Minimax Estimators of the Mean of a Multivariate Normal Distribution," Annals of Statistics, 4, 11-21.

——————— (1977), Comment on A Simulation Study of Alternatives to Ordinary Least Squares by A. P. Dempster, Martin Schatzoff, and Nanny Wermuth, 72, 91-3.

Farebrother, Richard W. (1975), "The Minimum Mean Square Error Linear Estimator and Ridge Regression," Technometrics, 17, 127-8.

Guilkey, David K. and Murphy, James L. (1975), "Directed Ridge Regression Techniques in Cases of Multicollinearity," Journal of the American Statistical Association, 70, 769-75.

Gunst, Richard F., and Mason, Robert L. (1977), "Mean Squared Error Properties of Regression Estimates," Journal of the American Statistical Association, 72, 616-28.

Hoaglin, David C. (1976), "Theoretical Properties of Congruential Random-Number Generators: An Empirical View," Memorandum NS-340, Department of Statistics, Harvard University.

Hocking, Ronald R. (1976), "The Analysis and Selection of Variables in Linear Regression," Biometrika, 32, 1-49.

Hoerl, Arthur E., and Kennard, Robert W. (1970a), "Ridge Regression: Biased Estimation for Non-Orthogonal Problems," Technometrics, 12, 55-67.

——————————(1970b), "Ridge Regression: Applications to Non-Orthogonal Problems," Technometrics, 12, 69-82.

—————————— and Baldwin, Kent F. (1975), "Ridge Regression: Some Simulations," Communications in Statistics, 4, 105-23.

International Business Machines (1970), System/360 Scientific Subroutine Package (360A-CM-03X), New York: IBM Technical Publications Department.

James, W., and Stein, Charles (1961), "Estimation with Quadratic Loss," Proceedings of the Fourth Berkeley Symposium on Mathmatical Statistics and Probability, University of California Press, 1, 361-79.

Lawless, Jerald F., and Wang, P. (1976) "A Simulation of Ridge and Other Regression Estimators," Communications in Statistics, A5, 307-23.

McDonald, Gary C., and Galarneau, Diane I. (1975), "A Monte Carlo Evaluation of Some Ridge-Type Estimators," Journal of the American Statistical Association, 70, 407-16.

Mallows, Colin L. (1973), "Some Comments on Cp," Technometrics, 15, 661-75.

Mitra, Amitava (1977), "A Comparison of Some Biased Estimators in Regression Analysis," Unpublished Ph.D. Dissertation, Department of Mathematical Sciences, Clemson University.

Newhouse, Joseph P., and Oman, Samuel D. (1971), "An Evaluation of Ridge Estimators," Report prepared for United States Air Force Project R-716-PR, Rand Corporation.

Vinod, Hrishikesh D. (1976), "A Ridge Estimator Whose MSE Dominates OLS," unpublished manuscript, Bell Telephone Laboratories, Holmdel, New Jersey.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER N-94	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Monte Carlo Comparison of Some Ridge and Other Biased Estimators		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER TR 272
7. AUTHOR(s) Amitava Mitra and Robert F. Ling		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0451
9. PERFORMING ORGANIZATION NAME AND ADDRESS Clemson University, Dept. of Mathematical Sciences, Clemson, South Carolina 29631		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-271
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, Code 436, Arlington, Va. 22217		12. REPORT DATE January, 1978
		13. NUMBER OF PAGES 32
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Ridge estimator; Biased estimator; Ordinary Least Squares; Multicollinearity; Squared loss; MSE.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A comparative study of the performance of 8 ridge estimators, 5 other biased estimators, and the Ordinary Least Squares estimate is made through a Monte Carlo experiment. The control parameters used in the simulation study are: (1) A measure of the degree of multicollinearity in X, (2) The number of independent variables, (3) The variance of the error term (for unit length parameter vector β), and (4) The most favorable and the least favorable choices of β (as defined by McDonald and Galarneau). 100 replications are

next page

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

beta

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. **ABSTRACT** (Continued)

run on each combination of $6 \times 3 \times 5 \times 2$ factorial design, using sample sizes of $n = 100$. The empirical mean squared error as well as some other properties of the sampling distributions of the squared loss $(\hat{\underline{\beta}} - \underline{\beta})'(\hat{\underline{\beta}} - \underline{\beta})$ are reported for the estimators considered.

beta-bar - beta